Comparing Two Populations

Dr. Jordan Srour

BUS 301: Int. Bus. Stats
Lebanese American University

6 November 2013

Notes

Today's Lecture

- Quiz 1 graded, average 81, will return it when all students have taken the exam.
- Assignment 4 assigned, Due 13 November
- Overview of Ch. 13
- Unpaired Test of Means, Equal Variance (13.1)
- Unpaired Test of Means, Unequal Variance (13.1)
- Testing Variance (13.4)

Announcement: We will extend class by 10 minutes each day (6, 8, 11, 13); No class on 15.

Notes

Assignment 4, due at the START of class, 13 November

1. 13.24 (Show your work)
2. 13.47
3. 13.55
4. 13.59 (You will need to use the data sets from blackboard for this.)
5. 13.81

Notes

Overview of Chapter 13

http://prezi.com/odc79llqtwhi/?utm_campaign=share&utm_medium=copy

Notes

Example 1: Two Versions of an Exam

When I give two versions of an exam (different in the order of questions), I randomly assign the students to the different versions and then carefully keep track of the scores for each questions. When all of the papers have been graded I check question by question to make sure that the order of the questions didn't give one group an unfair advantage over the other. Here is the data related to one of the questions on the recent exam.

	Mean	Std. Dev.	n
Exam Type 1	16.18	2.92	19
Exam Type 2	16.23	2.57	19

Given this data, what can I conclude? and how?

Notes

How can we compare the two sample means?

Subtract them! $\bar{x}_1 - \bar{x}_2$

How does this compare to the difference of the population means? It's expected value should be equal!

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$

Now we can make a test of the following hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$

Notes

What's the variance of $\bar{x}_1 - \bar{x}_2$ relative to the variance in the population?

How could we combine both variances?

$$V(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

So, now, what test statistic could we use?

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Great! Let's do a test. . . but wait! We don't have σ_1 or σ_2. What assumption can we make? The sample variances are a good approximation of the populations' variances! But how does that change our test statistic? We will need to combine the sample variances and we need to compute t.

Notes

How can we combine the sample standard deviations?

We still need one more assumption. . . The variance in population 1 EQUALS the variance in population 2. If the variances are equal we can use a weighted average! This is called the pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Notes

And the t-stat?

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

This t-stat can be used just as before except that now the degrees of freedom are $n_1 + n_2 - 2$.

Notes

Back to Example

In the example, t = -.05603; what can we conclude?

Notes

Example 2: Two Versions of an Exam

When I give two versions of an exam (different in the order of questions), I randomly assign the students to the different versions and then carefully keep track of the scores for each questions. When all of the papers have been graded I check question by question to make sure that the order of the questions didn't give one group an unfair advantage over the other. Here is the data related to another one of the questions on the recent exam.

	Mean	Std. Dev.	n
Exam Type 1	26.74	7.09	19
Exam Type 2	30.39	4.24	19

Given this data, what can I conclude? and how? Use a *t*-test!

Notes

Step 1. Hypotheses

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$

Notes

Step 2. Calculate t

If the variances are unequal, what is t?
Recall:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

What assumption can we make? The sample variances are a good approximation of the populations' variances! But now that we are assuming that the variances are unequal, the replacement is easier...

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Notes

Step 3. Look up $t_{\alpha/2}$

How many degrees of freedom should I use?

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Notes

Example 2:

$$t = -1.93,\ t_{.025,29} = 2.05,\ -t_{.025,29} = -2.05$$

What can we conclude? Fail to reject, H_0. The means for this exam question are equal.

Notes

How can we tell if the variances are equal?

What mathematical operators do we have to compare two values? Subtraction and division!

This time we will use division.

$$H_0 : \sigma_1^2/\sigma_2^2 = 1$$
$$H_1 : \sigma_1^2/\sigma_2^2 \neq 1$$

What test should we use? F-test! This is actually the topic of section 13.4. Test Statistic:

$$F = \frac{s_1^2}{s_2^2}$$

F-distributed with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.
`http://mips.stanford.edu/courses/stats_data_analsys/lesson_5/234_8_m.html`

Notes

Exercise 3: Comparing the total exam score between types

After comparing question by question, I also compare across the total exam scores. Here's the data, work with a partner and tell me if the exam was fair across the two different versions.

	Mean	Std. Dev.	n
Exam Type 1	77.97	16.12	19
Exam Type 2	84.14	13.79	19

Note, $F_{\alpha/2,\nu_1,\nu_2} = F_{.025,18,18} = 2.596$ and $F_{1-\alpha/2,\nu_1,\nu_2} = F_{.975,18,18} = .385$
HELPFUL EQUATIONS:

Equal variance:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$
$$\nu = n_1 + n_2 - 2.$$

Unequal variance:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}.$$

Notes

Notes